Graph-Adaptive Semi-Supervised Tracking of Dynamic Processes Over Switching Network Modes

Qin Lu, *Member, IEEE*, Vassilis N. Ioannidis, *Student Member, IEEE*, and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—A plethora of network-science related applications call for inference of spatio-temporal graph processes. Such an inference task can be aided by the underlying graph topology that might jump over discrete modes. For example, the connectivity in dynamic brain networks, switches among candidate topologies, each corresponding to a different emotional state, also known as the network mode. Taking advantage of limited nodal observations, the present contribution deals with semi-supervised tracking of dynamic processes over a given candidate set of graphs with *unknown* switches. Towards this end, a dynamical model is introduced to capture the per-slot spatial correlation using the active topology, as well as the temporal variation across slots through a state-space model. A scalable graph-adaptive Bayesian approach is developed, based on what is termed interacting multi-graph model (IMGM), to track the dynamic nodal processes and the active graph topology on-the-fly. Besides switching topologies, the proposed IMGM algorithm can accommodate various generalizations, including multiple dynamic functions, multiple kernels, and adaptive observation noise covariances. IMGM learns the dynamical model that best fits the data from a pool of available models. Thus, the resultant adaptive algorithm does not require offline model training. Numerical tests with synthetic and real datasets demonstrate the superior tracking performance of the novel approach compared to the mode-clairvoyant existing alternatives.**

***Index Terms*—Dynamic graph processes, switching network modes, online scalable Bayesian inference, multi-kernel learning.**

I. Introduction

GRAPHS capture relations among entities (nodes), and have found widespread application in various fields, including sociology, biology, neuroscience and economics [15], [30]. Attributes collected in interdependent feature vectors per node represent processes over the graph. Given such vectors from a subset of nodes, various applications call for semi-supervised learning (SSL) of processes across all network nodes. The scarcity of nodal observations can be due to e.g., cost, and computational or privacy constraints. For example, individuals in social networks may be reluctant to share personal information, while acquiring nodal samples in brain networks may require invasive procedures such as electrocorticography.

Manuscript received June 1, 2019; revised January 31, 2020 and March 16, 2020; accepted March 23, 2020. Date of publication April 3, 2020; date of current version May 1, 2020. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Pierre Borgnat. This work was supported by NSF under Grants 1508993, 1711471, and 1901134. *(Corresponding author: Qin Lu.)*

The authors are with the Department of ECE and Digital Technology Center, University of Minnesota, Minneapolis, MN 55414 USA (e-mail: luqinrosy@gmail.com; ioann006@umn.edu; georgios@umn.edu).

Digital Object Identifier 10.1109/TSP.2020.2984889

SSL tasks over networks can leverage the prior information of the underlying graph topology that captures nodal interdependencies [12]. Existing approaches to reconstructing *time-invariant* (TI) graph processes often rely on the smoothness of graph processes [16], [31], which asserts that connected vertices have similar features. In social networks where nodes and edges represent users and their friendships, one can infer the age of a specific user from her or his friends' age. Other than smoothness inference from limited nodal observations can rely on e.g., 'graph bandlimitedness' [10], [29], sparsity and overcomplete dictionaries [11]. Most of these approaches can be unified under the framework of learning using graph kernels; see e.g., [26].

The aforementioned SSL task becomes more challenging when nodal processes are *nonstationary*, and the graph topology is also *time-varying*. In a brain network for instance, where nodes correspond to brain regions and edges capture dependencies among them, one may be interested in predicting the dynamic processes as well as the varying interconnections. An interesting time-varying topology model *switches* over a set of connectivity patterns, also known as "network modes" [5]. For example, the connectivity among human brain regions varies as the humans' emotional, mental or physical activities change [36]. Coupled with the topology, the dynamics of nodal processes can also switch among different modes. Switching dynamical models have been typically employed to characterize the multi-modal behavior of control systems [27], as well as kinematics of maneuvering targets such as drones [6]. Nevertheless, graph-based switching dynamical models have not been considered so far.

Several attempts have been made to reconstruct dynamic graph processes in the presence of possibly time-varying topologies. Inference of slow-varying processes over graphs has been pursued using the so-termed graph bandlimited model in [10], [35]. On the other hand, graph kernel-based estimators have been leveraged in [25], [14] to reconstruct general dynamic processes. All these contemporary approaches rely on a *known* graph topology and *fixed* dynamic models. However, the dynamic graph can change or switch in an *unknown* fashion among a set of possibly known topologies, which may reflect sudden changes in the partially observed signals. Furthermore, even when no topology switches occur, the graph process can evolve over *multiple* dynamical models across time, and thus a fixed model may be inadequate.

The present paper puts forth an approach for semi-supervised tracking and extrapolation of dynamic nodal processes over switching graphs. Our contribution is threefold.

1053-587X © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: University of Minnesota. Downloaded on July 09,2021 at 01:19:13 UTC from IEEE Xplore. Restrictions apply.

C1. The evolution of dynamic processes over switching graphs is captured by a first-order vector autoregressive model, where the transition matrix and the process noise covariance matrix depend on the active mode-conditioned topology. The resulting graph-adaptive dynamical model accounts for both spatial correlation within one slot and temporal variations across slots.

C2. Given a candidate set of the aforementioned mode-conditioned dynamical models and measurements on a subset of nodes, we put forth a scalable graph-aware Bayesian tracker, termed interacting multi-graph model (IMGM), to jointly estimate the graph processes and active network modes on-the-fly.

C3. Further, the proposed IMGM framework accommodates various modeling extensions, including switching nonlinear dynamical functions, multiple kernels, and adaptive observation covariances. By accounting for these dynamical models, IMGM adapts on the observed data and selects the pertinent model per time slot without requiring offline training.

If observations were available at all nodes, it would have been possible to identify the active topology per slot without explicitly modeling the nodal process dynamics [5]. Relative to [5], this work leverages the dynamics to reconstruct unavailable nodal data, while at the same time identifying the active mode and tracking the nodal processes. Not necessarily graph related yet similar to that of [5] is the goal of subspace clustering [32], but different from the work here mode dynamics are not exploited to reconstruct unavailable nodal processes.

The rest of the paper is organized as follows. Section II starts with preliminaries to formulate the problem that is solved in Section III. Section IV deals with modeling generalizations of the IMGM approach. Numerical results and conclusions are presented in Sections V and VI, respectively. Part of this paper is published in our conference precursors [20], [21].

Notation: Scalars are denoted by lowercase, column vectors by bold lowercase, and matrices by bold uppercase fonts. Superscripts $^\top$, $^{-1}$ and † denote transpose, inverse and pseudo-inverse, respectively; while $\mathbf{1}_N$ stands for the $N \times 1$ all-ones vector; and $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \mathbf{K})$ for the probability density function (pdf) of a Gaussian random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$, and covariance matrix $\mathbf{K}$. Finally, if $\mathbf{A}$ is a matrix and $\mathbf{x}$ a vector, then $||\mathbf{x}||^2_{\mathbf{A}} := \mathbf{x}^\top \mathbf{A}^{-1}\mathbf{x}$, $||\mathbf{x}||^2_2 := \mathbf{x}^\top\mathbf{x}$, $\|\mathbf{A}\|_1$ represents the $\mathcal{L}_1$-norm of the vectorized matrix, and $\|\mathbf{A}\|^2_F$ is the Frobenius norm of $\mathbf{A}$.

II. Problem Formulation

Consider a time-varying graph $\mathcal{G}_t$ with N nodes indexed by the vertex set $\mathcal{V} := \{1, \ldots, N\}$. Per slot t, the relationship between nodes is captured by an $N \times N$ adjacency matrix $\mathbf{A}_t$ with $A_t(n, n')$ representing the weight of the edge connecting nodes n and n'. The focus will be on graphs whose topology jumps among a *known* set of S candidate adjacency matrices; that is, $\mathbf{A}_t = \mathbf{A}_t^{\sigma_t} \in \{\mathbf{A}_t^1, \ldots, \mathbf{A}_t^S\}$, where the per-slot active topology index $\sigma_t \in \mathcal{S} := \{1, \ldots, S\}$ describes the so-called "network mode." The active mode-conditioned Laplacian matrix is then given by $\mathbf{L}_t^{\sigma_t} = \mathbf{D}_t^{\sigma_t} - \mathbf{A}_t^{\sigma_t}$, where $\mathbf{D}_t^{\sigma_t} = \mathrm{diag}\{\mathbf{A}_t^{\sigma_t}\mathbf{1}_N\}$ denotes the graph degree matrix. Switching topologies emerge in several networked systems. Besides brain networks [36], network topologies from information cascades exhibit switching patterns [5].

TABLE I
Examples of Laplacian Kernels

Kernel type	Function	Parameters
Diffusion	$r(\lambda) = \exp\{a^2\lambda/2\}$	$a^2 \geq 0$
Regularized Laplacian	$r(\lambda) = 1 + a^2\lambda$	$a^2 \geq 0$
Bandlimited	$r(\lambda_n) = \begin{cases} 1/\beta & 1 \leq n \leq B \\ \beta & \text{otherwise} \end{cases}$	$\beta \gg 1, B$
p-step random walk	$r(\lambda) = (a - \lambda)^{-p}$	a, p

A dynamic graph process is defined as the mapping $x : \mathcal{V} \times \mathcal{T} \mapsto \mathbb{R}$, where $\mathcal{T} := \{1, 2, \ldots\}$ is the set of slot indices. Thus, $x_t(n)$ represents the attribute of node n at slot t. For instance, it may represent the value of a stock n at day t. The values over all the nodes at slot t are collected in the vector $\mathbf{x}_t := [x_t(1), \ldots, x_t(N)]^\top$.

In several applications, processes over only a subset of $M < N$ vertices are observed, which yields the observation model

$$\mathbf{z}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{e}_t \tag{1}$$

where $\mathbf{H}_t \in \{0, 1\}^{M \times N}$ is the time-varying observation (or sampling) matrix, whose rows sum up to 1, and $\mathbf{e}_t$ is the observation noise that accounts for unmodeled uncertainties, assumed to be white and Gaussian distributed with mean zero and covariance $\mathbf{R}_t$.

A. Kernel-Based Inference of TI Graph Processes

Towards learning dynamic graph processes, it is instructive to first outline the kernel-based inference of TI graph processes. Consider a TI adjacency matrix $\mathbf{A}$ and observation model $\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{e}$, which are given by dropping slot index t in the time-varying scenario. To uniquely reconstruct $\mathbf{x}$, one may rely on the regularized least-squares formulation

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{z} - \mathbf{H}\mathbf{x}\|_2^2 + \mu\Omega(\mathbf{x}) \tag{2}$$

where $\Omega(\cdot)$ is a chosen monotonic regularizing function along with the scalar $\mu \geq 0$ that controls the importance of the regularization term vis-a-vis the fitting error.

For undirected graphs with symmetric adjacency matrix $\mathbf{A}$, the so-called *Laplacian regularizer* is given by

$$\Omega_{\mathrm{LR}}(\mathbf{x}) := \mathbf{x}^\top\mathbf{L}\mathbf{x} = \frac{1}{2}\sum_{n=1}^{N}\sum_{n'=1}^{N} A(n, n')(x(n) - x(n'))^2 \tag{3}$$

where $\mathbf{L}$ is the TI Laplacian matrix. The regularizer (3) promotes smoothness of the estimated signal on the graph as vertices connected by strong links (large $A(n, n')$) will have similar signal estimates to minimize (3). To facilitate other properties such as diffusion or graph bandlimitedness, the Laplacian matrix in (3) is replaced by $r(\mathbf{L})$, where the scalar energy mapping $r : \mathbb{R} \mapsto \mathbb{R}_+$ is applied on the eigenvalues of $\mathbf{L}$ to promote desired properties, see e.g., Table I. The pseudo-inverse of $r(\mathbf{L})$

Authorized licensed use limited to: University of Minnesota. Downloaded on July 09,2021 at 01:19:13 UTC from IEEE Xplore. Restrictions apply.

yields the graph Laplacian kernel [26]

$$\mathbf{K} := r^{\dagger}(\mathbf{L}). \tag{4}$$

By considering $\Omega(\mathbf{x}) := \|\mathbf{x}\|^2_{\mathbf{K}}$, we recover the family of *kernel ridge regression* (KRR) estimators, which enjoys well-documented reconstruction performance [14], [31].

For directed graphs, one can not directly apply the KRR framework since $\mathbf{A}$ is not symmetric. Nevertheless, attempts have also been made towards KRR by redefining a positive semidefinite Laplacian matrix [3], [4], [8], [9]. In [9], such a valid matrix is constructed as $\mathbf{L} := \mathbf{U} - (\mathbf{U}\bar{\mathbf{A}} + \bar{\mathbf{A}}^{\top}\mathbf{U})/2$, where $\bar{\mathbf{A}} := \mathbf{D}^{-1}\mathbf{A}$ and $\mathbf{U} := \mathrm{diag}(\mathbf{u})$ with $\mathbf{u}$ denoting the left eigenvector of $\bar{\mathbf{A}}$. Based on this definition, kernel matrices can be constructed, allowing the KRR framework to accommodate directed graphs as well. The methods in this paper apply to both directed and undirected graphs.

So far, we outlined SSL on graphs using a deterministic kernel-based framework. It is however instructive to present a Bayesian generative model for KRR estimation. First, consider that the prior pdf of $\mathbf{x}$ is $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \mathbf{0}, \mathbf{K})$ and the likelihood of $\mathbf{x}$ based on observation $\mathbf{z}$ is given by $p(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{z}; \mathbf{H}\mathbf{x}, \mathbf{R})$ with $\mathbf{R} = \mu\mathbf{I}_M$. Under these Gaussian densities, the maximum a posteriori (MAP) estimator of $\mathbf{x}$ given $\mathbf{z}$ is equivalent to the KRR estimator, which amounts to the linear minimum mean-square error (LMMSE) estimator

$$\begin{aligned}\hat{\mathbf{x}} &= \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{z}) = \arg\max_{\mathbf{x}} p(\mathbf{z}|\mathbf{x})p(\mathbf{x}) \\ &= \arg\min_{\mathbf{x}} \|\mathbf{z} - \mathbf{H}\mathbf{x}\|^2_{\mathbf{R}} + \|\mathbf{x}\|^2_{\mathbf{K}}. \end{aligned} \tag{5}$$

Graph processes with arbitrary dynamics render the inference task in (5) intractable, in general. Fortunately, structured dynamical models, such as the one dealt with in the ensuing section, can lead to tractable estimators.

B. Modeling Dynamic Processes Over Switching Graphs

One possible approach is to pursue an instantaneous per-slot KRR estimator based on $\mathbf{z}_t$ in (1). This estimator however, does not account for the $\mathbf{x}_{t-1}$ to $\mathbf{x}_t$ transition that can benefit the estimator of $\mathbf{x}_t$ from observations other than $\mathbf{z}_t$, and thus improve estimation performance [14], [25].

Exploitation of graph process dynamics calls for modeling the evolution from $\mathbf{x}_{t-1}$ to $\mathbf{x}_t$, which arguably depends on the underlying topology [14], [25]. To capture the dynamics of processes over switching graphs, we model the evolution from $\mathbf{x}_{t-1}$ to $\mathbf{x}_t$ as the first-order Markov process

$$\mathbf{x}_t = \mathbf{F}_t^{\sigma_t}\mathbf{x}_{t-1} + \boldsymbol{\eta}_t^{\sigma_t} \tag{6}$$

where the state transition matrix is a known function f of the active adjacency matrix given by

$$\mathbf{F}_t^{\sigma_t} := f\left(\mathbf{A}_t^{\sigma_t}\right). \tag{7}$$

The mode-conditioned process noise $\boldsymbol{\eta}_t^{\sigma_t}$ is assumed uncorrelated with the state, white, and Gaussian distributed with zero mean and covariance $\mathbf{K}_t^{\sigma_t}$, which is the Laplacian kernel (4).

The model in (6) accounts for the spatio-temporal dependence of graph processes in the following two aspects.

i) The temporal dynamics across two consecutive slots are captured by the state transition matrix of the so-termed "transition graph." With $\mathbf{F}_t^{\sigma_t} = \mathbf{A}_t^{\sigma_t}$, the transition model (6) amounts to a graph diffusion process [29].
ii) The spatial correlations across nodes within t are captured by the Laplacian kernel $\mathbf{K}_t^{\sigma_t}$ of the process noise covariance. By setting $\mathbf{F}_t^{\sigma_t}\mathbf{x}_{t-1} = \mathbf{0}$, the dynamical model (6) reduces to $\mathbf{x}_t = \boldsymbol{\eta}_t^{\sigma_t}$, which together with (1), constitutes the generative model for TI graph processes, leading to the MAP estimate given by (5). Incidentally, such a covariance model implies that $\mathbf{x}_t$ is "graph stationary" [24]. A related noise model was also adopted in [14] to promote smoothness of the estimates.

The dynamical model in (6) describes what is also known as a switching linear dynamical system (SLDS) [23], and it is widely employed in the tracking community to capture the kinematic state evolution of maneuvering targets [6].

Problem statement: Given T observations $\mathbf{Z}_T := [\mathbf{z}_1 \dots \mathbf{z}_T]$ as in (1), and candidate models $\{\{\mathbf{F}_t^s, \mathbf{K}_t^s\}_{s=1}^S\}_{t=1}^T$ as in (6), the goal is to jointly track the dynamic graph processes $\mathbf{X}_T := [\mathbf{x}_1 \dots \mathbf{x}_T]$, and the discrete modes $\{\sigma_t\}_{t=1}^T$.

III. Scalable Graph-Aware Bayesian Tracker

In this section, we develop a Bayesian approach to track dynamic graph processes over switching graphs. First, given the Markovian state transition model in (6), the prior joint pdf of the nodal processes in $\mathbf{X}_T$ can be expressed as

$$\begin{aligned} p(\mathbf{X}_T) &= p(\mathbf{x}_T|\mathbf{x}_{T-1}; \sigma_T)p(\mathbf{X}_{T-1}) = \cdots = \prod_{t=1}^{T} p(\mathbf{x}_t|\mathbf{x}_{t-1}; \sigma_t) \\ &= \prod_{t=1}^{T}\left(\sum_{s=1}^{S} w_t^s p(\mathbf{x}_t|\mathbf{x}_{t-1}; \sigma_t = s)\right) \end{aligned}$$

where we explicitly incorporate σ_t in $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ to stress the active topology present, and w_t^s encodes the existence of the mode $\sigma_t = s$ with $w_t^s \in \{0, 1\}$ and $\sum_{s=1}^S w_t^s = 1$.

Furthermore, since $\mathbf{e}_t$ in (1) is temporally white, the conditional data pdf also factorizes as

$$p(\mathbf{Z}_T|\mathbf{X}_T) = \prod_{t=1}^{T} p(\mathbf{z}_t|\mathbf{x}_t).$$

Hence, Bayes' rule yields the posterior joint state pdf as

$$\begin{aligned} p(\mathbf{X}_T|\mathbf{Z}_T) &\propto p(\mathbf{Z}_T|\mathbf{X}_T)p(\mathbf{X}_T) \\ &= \prod_{t=1}^{T} p(\mathbf{z}_t|\mathbf{x}_t)\left(\sum_{s=1}^{S} w_t^s p(\mathbf{x}_t|\mathbf{x}_{t-1}; \sigma_t = s)\right). \end{aligned} \tag{8}$$

Since, $\mathbf{e}_t$ and $\boldsymbol{\eta}_t^{\sigma_t}$ are Gaussian, the conditional likelihood $p(\mathbf{z}_t|\mathbf{x}_t)$ and the transition pdf $p(\mathbf{x}_t|\mathbf{x}_{t-1}; \sigma_t = s)$ are also Gaussian, that is

$$\begin{aligned} p(\mathbf{z}_t|\mathbf{x}_t) &= \mathcal{N}(\mathbf{z}_t; \mathbf{H}_t\mathbf{x}_t, \mathbf{R}_t) \\ p(\mathbf{x}_t|\mathbf{x}_{t-1}; \sigma_t = s) &= \mathcal{N}(\mathbf{x}_t; \mathbf{F}_t^s\mathbf{x}_{t-1}, \mathbf{K}_t^s). \end{aligned}$$

Authorized licensed use limited to: University of Minnesota. Downloaded on July 09,2021 at 01:19:13 UTC from IEEE Xplore. Restrictions apply.

Thus, the MAP state estimates in batch form are (cf. (8))

$$\underset{\substack{\{\mathbf{x}_t\}_{t=1}^T \\ \{\{w_t^s\}_{s=1}^S\}_{t=1}^T}}{\arg\min} \frac{1}{2}\sum_{t=1}^{T}\left[\|\mathbf{z}_t-\mathbf{H}_t\mathbf{x}_t\|^2_{\mathbf{R}_t}+\sum_{s=1}^{S} w_t^s\|\mathbf{x}_t-\mathbf{F}_t^s\mathbf{x}_{t-1}\|^2_{\mathbf{K}_t^s}\right]$$

$$\text{s.to}\quad w_t^s\in\{0,1\},\quad \sum_{s=1}^{S} w_t^s=1. \tag{9}$$

Unfortunately, (9) is a mixed integer program whose optimal solution is given by enumerating all the S^T combinations of discrete network modes across T slots, and then applying a Kalman smoother for each mode combination, thus incurring computational complexity $\mathcal{O}(TS^TN^3)$.

Targeting a computationally efficient solver with $\mathbf{x}_t$ and σ_t estimates obtained on-the-fly, we will build on the interacting multi-model (IMM) algorithm [7] that has been applied to target tracking [22] and air traffic control [19], but without graph-related information. Taking into account dynamically switching graph topologies, we will naturally term the resultant algorithm interacting multi-graph model (IMGM). Given partially observed nodal samples $\mathbf{z}_t$ and a candidate set of switching graphs, IMGM is a *graph-adaptive* Bayesian tracker that estimates the active network mode σ_t together with the N scalar nodal processes in $\mathbf{x}_t$.

Our IMGM replaces the hard constraint $w_t^s\in\{0,1\}$ with the soft one $w_t^s\in[0,1]$. To further stress that the weight is based on observations up to t, w_t^s is replaced with $w_{t|t}^s$. Thus, one can interpret $w_{t|t}^s$ as the posterior probability mass function (pmf) of mode s being active at slot t, namely $w_{t|t}^s=\Pr(\sigma_t=s|\mathbf{Z}_t)$. Different from (9) where σ_t was viewed as deterministic, we will next model it as a first-order Markov chain parameterized by the $S\times S$ mode transition matrix $\mathbf{\Pi}$, whose (i,j)th entry

$$\pi_{ij}=\Pr(\sigma_t=i|\sigma_{t-1}=j) \tag{10}$$

denotes the transition probability from mode j at slot $t-1$ to mode i at slot t. The parameters of $\mathbf{\Pi}$ are pre-selected. A practical choice for $\mathbf{\Pi}$ is to set its diagonal entries to $\pi_0\in[0.9,1)$, and the rest to $(1-\pi_0)/(S-1)$ [6].

IMGM leverages the current observation $\mathbf{z}_t$ to propagate the posterior marginal state pdf $p(\mathbf{x}_{t-1}|\mathbf{Z}_{t-1})$ to $p(\mathbf{x}_t|\mathbf{Z}_t)$. Towards this end, we start by approximating the mode-conditional posterior of $\mathbf{x}_t$ with a Gaussian pdf

$$p(\mathbf{x}_t|\sigma_t=s,\mathbf{Z}_t)\approx\mathcal{N}(\mathbf{x}_t;\hat{\mathbf{x}}_{t|t}^s,\mathbf{P}_{t|t}^s) \tag{11}$$

where $\hat{\mathbf{x}}_{t|t}^s$ and $\mathbf{P}_{t|t}^s$ are the mean and the covariance matrix associated with mode s. Bayes' rule and the total probability theorem (TPT) yields the marginal posterior

$$\begin{aligned} p(\mathbf{x}_t|\mathbf{Z}_t)&=\sum_{s=1}^{S}\Pr(\sigma_t=s|\mathbf{Z}_t)\,p(\mathbf{x}_t|\sigma_t=s,\mathbf{Z}_t)\\ &\approx\sum_{s=1}^{S} w_{t|t}^s\,\mathcal{N}(\mathbf{x}_t;\hat{\mathbf{x}}_{t|t}^s,\mathbf{P}_{t|t}^s) \end{aligned} \tag{12}$$

approximated by a Gaussian mixture (GM) pdf, which is parameterized by the set

$$\mathcal{P}_t:=\{w_{t|t}^s,\hat{\mathbf{x}}_{t|t}^s,\mathbf{P}_{t|t}^s,s=1,\ldots,S\}. \tag{13}$$

This GM model facilitates the propagation from $p(\mathbf{x}_{t-1}|\mathbf{Z}_{t-1})$ to $p(\mathbf{x}_t|\mathbf{Z}_t)$ through updates of the elements in $\mathcal{P}_{t-1}$ to those in $\mathcal{P}_t$. These updates will be implemented using the prediction and correction of the mode pmf and the mode-conditional state pdf as detailed next.

A. Prediction

At the end of slot $t-1$, the posterior marginal state pdf is characterized by $\mathcal{P}_{t-1}$. Before the arrival of a new observation $\mathbf{z}_t$, IMGM leverages the mode and state evolution models (cf. (10) and (6)) to make predictions about the mode pmf and the mode-conditional state pdf, respectively.

1) Predicted Mode Pmf: Based on the Markov transition model (10), the predicted mode pmf is readily obtained via TPT and Bayes' rule as

$$\begin{aligned} w_{t|t-1}^s&:=\Pr(\sigma_t=s|\mathbf{Z}_{t-1})=\sum_{s'=1}^{S}\Pr(\sigma_t=s,\sigma_{t-1}=s'|\mathbf{Z}_{t-1})\\ &=\sum_{s'=1}^{S}\Pr(\sigma_t=s|\sigma_{t-1}=s,'\,\mathbf{Z}_{t-1})\Pr(\sigma_{t-1}=s'|\mathbf{Z}_{t-1})\\ &=\sum_{s'=1}^{S}\pi_{ss'}w_{t-1|t-1}^{s'}\,. \end{aligned} \tag{14}$$

2) Predicted State Pdf: Since $p(\mathbf{x}_{t-1}|\sigma_{t-1}=s,'\,\mathbf{Z}_{t-1})$ is Gaussian (cf. (11)), the linear-Gaussian state transition model (6) conditioned on $\sigma_t=s$ allows one to deduce that

$$p(\mathbf{x}_t|\sigma_t=s,\sigma_{t-1}=s,'\,\mathbf{Z}_{t-1})=\mathcal{N}(\mathbf{x}_t;\hat{\mathbf{x}}_{t|t-1}^{s,s'},\mathbf{P}_{t|t-1}^{s,s'}) \tag{15}$$

where the subscript (s,s') refers to the conditioning on modes (s,s') at slots t and $t-1$ respectively.

The first two moments of the pdf in (15) are given by

$$\hat{\mathbf{x}}_{t|t-1}^{s,s'}=\mathbf{F}_t^s\hat{\mathbf{x}}_{t-1|t-1}^{s'} \tag{16a}$$

$$\mathbf{P}_{t|t-1}^{s,s'}=\mathbf{F}_t^s\mathbf{P}_{t-1|t-1}^{s'}(\mathbf{F}_t^s)^\top+\mathbf{K}_t^s\,. \tag{16b}$$

Next, using the TPT and Bayes' rule, we express the predicted mode-conditional state pdf at t as

$$\begin{aligned} &p(\mathbf{x}_t|\sigma_t=s,\mathbf{Z}_{t-1})\\ &=\sum_{s'=1}^{S}\Pr(\sigma_{t-1}=s'|\sigma_t=s,\mathbf{Z}_{t-1})p(\mathbf{x}_t|\sigma_t=s,\sigma_{t-1}=s,'\,\mathbf{Z}_{t-1}) \end{aligned} \tag{17}$$

where $\Pr(\sigma_{t-1}=s'|\sigma_t=s,\mathbf{Z}_{t-1}):=w_{t-1|t}^{s'|s}$ can be interpreted as the backward mode transition probability, which upon

Authorized licensed use limited to: University of Minnesota. Downloaded on July 09,2021 at 01:19:13 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1: One Recursion of the IMGM Algorithm.

Input: $\mathcal{P}_{t-1}$, $\mathbf{z}_t$, $\{\mathbf{F}_t^s, \mathbf{K}_t^s\}_{s=1}^S$, $\mathbf{R}_t$, $\mathbf{H}_t$, $\mathbf{\Pi}$
for $s = 1$ to S **do**
  S1 Prediction
    S1.1 of mode pmf via (14)
    S1.2 of mode-conditional state pdf via (16), (18), and (20)
  S2 Correction
    S2.1 of mode-conditional state pdf via (23)
    S2.2 of mode pmf via (24)
  S3 Fusion of mode-conditional state pdfs via (27)
end for
Output: $\mathcal{P}_t$, $\hat{\mathbf{x}}_{t|t}$, $\mathbf{P}_{t|t}$

appealing to Bayes' rule and the TPT, boils down to

$$w_{t-1|t}^{s'|s} = \frac{\Pr(\sigma_{t-1}=s'|\mathbf{Z}_{t-1})\Pr(\sigma_t=s|\sigma_{t-1}=s,'\,\mathbf{Z}_{t-1})}{\sum_{s'=1}^S \Pr(\sigma_{t-1}=s'|\mathbf{Z}_{t-1})\Pr(\sigma_t=s|\sigma_{t-1}=s,'\,\mathbf{Z}_{t-1})} = \frac{w_{t-1|t-1}^{s'}\pi_{ss'}}{\sum_{s'=1}^S w_{t-1|t-1}^{s'}\pi_{ss'}}\,. \tag{18}$$

So far, the predicted mode-conditional state pdf (17) is a GM pdf. A GM prior however, does not evolve to a Gaussian posterior pdf with Gaussian likelihood. To maintain Gaussianity of the posterior mode-conditional state pdf as in (11), we will approximate (17) by the following Gaussian pdf

$$p(\mathbf{x}_t|\sigma_t=s,\mathbf{Z}_{t-1}) \approx \mathcal{N}(\mathbf{x}_t;\hat{\mathbf{x}}_{t|t-1}^s,\mathbf{P}_{t|t-1}^s) \tag{19}$$

where $\hat{\mathbf{x}}_{t|t-1}^s$ and $\mathbf{P}_{t|t-1}^s$ are chosen to match the first two moments of the GM pdf (17) as

$$\hat{\mathbf{x}}_{t|t-1}^s = \sum_{s'=1}^S w_{t-1|t}^{s'|s}\hat{\mathbf{x}}_{t|t-1}^{s,s'} \tag{20a}$$

$$\mathbf{P}_{t|t-1}^s = \sum_{s'=1}^S w_{t-1|t}^{s'|s}\Big(\mathbf{P}_{t|t-1}^{s,s'} + (\hat{\mathbf{x}}_{t|t-1}^{s,s'}-\hat{\mathbf{x}}_{t|t-1}^s)(\hat{\mathbf{x}}_{t|t-1}^{s,s'}-\hat{\mathbf{x}}_{t|t-1}^s)^\top\Big)\,. \tag{20b}$$

Approximating non-Gaussian pdfs with Gaussian ones is a well-documented approach to effect scalability in approximate (Bayesian) inference, including variational inference and expectation propagation; see [23] and the references therein. With moments of the approximating Gaussian pdf matched to that of the non-Gaussian one (cf. (20)), the KL divergence between the two pdfs is minimized.

Up to now, we have obtained the predicted mode pmf and the mode-conditional state pdf, which will be propagated to their posterior counterparts after a new $\mathbf{z}_t$ is observed.

B. Correction

1) Posterior Mode-Conditional State Pdf: Given the new observation $\mathbf{z}_t$, the approximate predicted mode-conditional state pdf (19) is propagated to its posterior via Bayes' rule as

$$p(\mathbf{x}_t|\sigma_t=s,\mathbf{Z}_t) = p(\mathbf{x}_t|\sigma_t=s,\mathbf{z}_t,\mathbf{Z}_{t-1}) = \frac{p(\mathbf{x}_t|\sigma_t=s,\mathbf{Z}_{t-1})p(\mathbf{z}_t|\mathbf{x}_t,\sigma_t=s,\mathbf{Z}_{t-1})}{p(\mathbf{z}_t|\sigma_t=s,\mathbf{Z}_{t-1})} \tag{21}$$

where $p(\mathbf{z}_t|\mathbf{x}_t,\sigma_t=s,\mathbf{Z}_{t-1}) = \mathcal{N}(\mathbf{z}_t;\mathbf{H}_t\mathbf{x}_t,\mathbf{R}_t)$, since $\mathbf{z}_t$ is independent of $\mathbf{Z}_{t-1}$ and σ_t. Hence, with the likelihood and the prior (cf. (19)) being Gaussian, it holds that

$$p(\mathbf{x}_t|\sigma_t=s,\mathbf{Z}_t) = \mathcal{N}(\mathbf{x}_t;\hat{\mathbf{x}}_{t|t}^s,\mathbf{P}_{t|t}^s) \tag{22}$$

where the first two moments $\hat{\mathbf{x}}_{t|t}^s$ and $\mathbf{P}_{t|t}^s$ are obtained via the Kalman update as (see e.g., [6])

$$\hat{\mathbf{z}}_{t|t-1}^s = \mathbf{H}_t\hat{\mathbf{x}}_{t|t-1}^s \tag{23a}$$

$$\mathbf{\Phi}_{t|t-1}^s = \mathbf{H}_t\mathbf{P}_{t|t-1}^s(\mathbf{H}_t)^\top + \mathbf{R}_t \tag{23b}$$

$$\mathbf{G}_t^s = \mathbf{P}_{t|t-1}^s(\mathbf{H}_t)^\top(\mathbf{\Phi}_t^s)^{-1} \tag{23c}$$

$$\hat{\mathbf{x}}_{t|t}^s = \hat{\mathbf{x}}_{t|t-1}^s + \mathbf{G}_t^s(\mathbf{z}_t-\hat{\mathbf{z}}_{t|t-1}^s) \tag{23d}$$

$$\mathbf{P}_{t|t}^s = \mathbf{P}_{t|t-1}^s - \mathbf{G}_t^s\mathbf{\Phi}_{t|t-1}^s(\mathbf{G}_t^s)^\top\,. \tag{23e}$$

2) Posterior Mode Pmf: Upon applying Bayes' rule, the posterior mode pmf is

$$w_{t|t}^s = \Pr(\sigma_t=s|\mathbf{z}_t,\mathbf{Z}_{t-1}) = \frac{p(\mathbf{z}_t|\sigma_t=s,\mathbf{Z}_{t-1})\Pr(\sigma_t=s|\mathbf{Z}_{t-1})}{\sum_{s=1}^S p(\mathbf{z}_t|\sigma_t=s,\mathbf{Z}_{t-1})\Pr(\sigma_t=s|\mathbf{Z}_{t-1})} \tag{24}$$

where the first factor $p(\mathbf{z}_t|\sigma_t=s,\mathbf{Z}_{t-1})$ is computable via (14), and the second factor is the normalizing pdf in (21)

$$p(\mathbf{z}_t|\sigma_t=s,\mathbf{Z}_{t-1}) = \int p(\mathbf{z}_t,\mathbf{x}_t|\sigma_t=s,\mathbf{Z}_{t-1})d\mathbf{x}_t = \int p(\mathbf{z}_t|\mathbf{x}_t)p(\mathbf{x}_t|\sigma_t=s,\mathbf{Z}_{t-1})d\mathbf{x}_t \tag{25}$$

that can be shown to be the Gaussian $\mathcal{N}(\mathbf{z}_t;\hat{\mathbf{z}}_{t|t-1}^s,\mathbf{\Phi}_{t|t-1}^s)$ with $\hat{\mathbf{z}}_{t|t-1}^s$ and $\mathbf{\Phi}_{t|t-1}^s$ given by (23a) and (23b), respectively.

C. Fusion

Finally, the marginal posterior state pdf is given by fusing the individual mode-conditional posteriors to obtain the GM

$$p(\mathbf{x}_t|\mathbf{Z}_t) = \sum_{s=1}^S w_{t|t}^s\mathcal{N}(\mathbf{x}_t;\hat{\mathbf{x}}_{t|t}^s,\mathbf{P}_{t|t}^s) \tag{26}$$

whose first two moments are

$$\hat{\mathbf{x}}_{t|t} = \sum_{s=1}^S w_{t|t}^s\hat{\mathbf{x}}_{t|t}^s \tag{27a}$$

$$\mathbf{P}_{t|t} = \sum_{s=1}^S w_{t|t}^s\Big(\mathbf{P}_{t|t}^s + (\hat{\mathbf{x}}_{t|t}^s-\hat{\mathbf{x}}_{t|t})(\hat{\mathbf{x}}_{t|t}^s-\hat{\mathbf{x}}_{t|t})^\top\Big)\,. \tag{27b}$$

Authorized licensed use limited to: University of Minnesota. Downloaded on July 09,2021 at 01:19:13 UTC from IEEE Xplore. Restrictions apply.

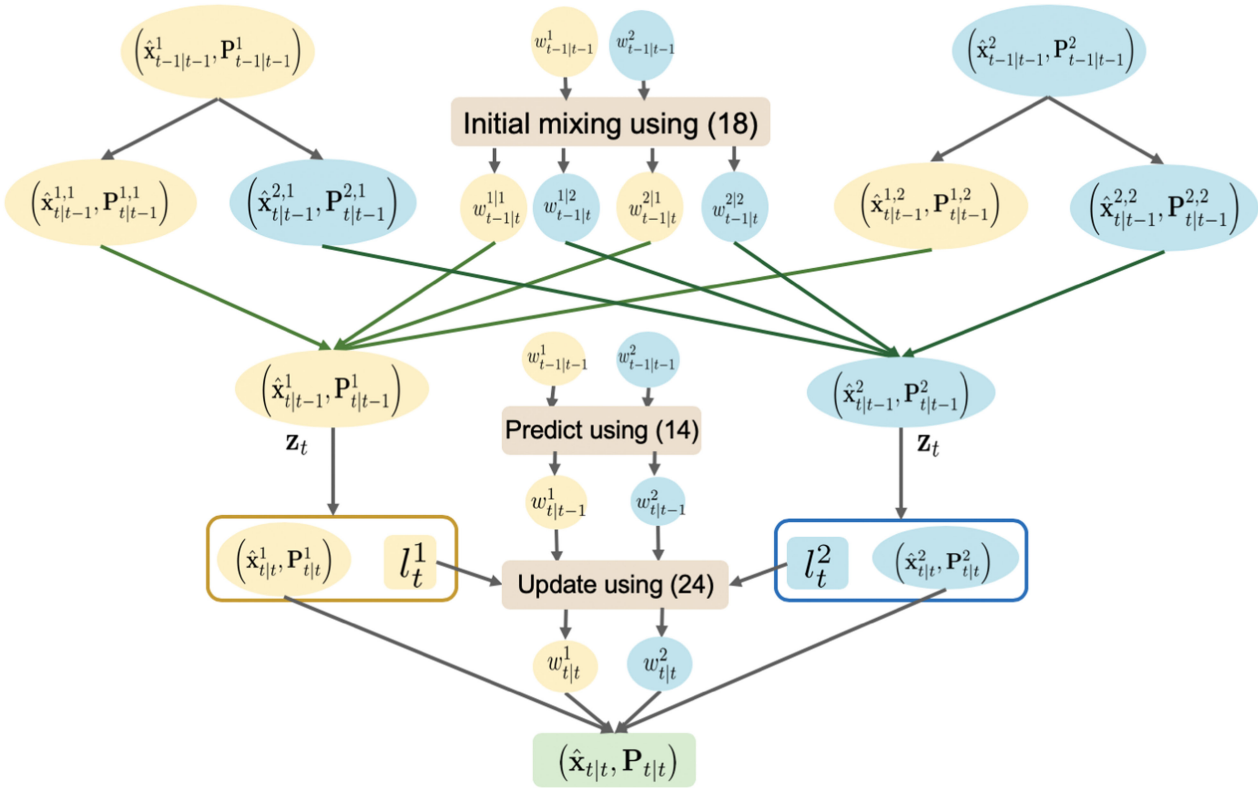


Fig. 1. Flowchart of IMGM with $S = 2$ modes for one recursion, where yellow is used for mode 1, and blue for mode 2. Each mode predicts the first two moments of the state pdf at slot t assuming the active mode is 1, or 2, respectively. Then the predictive state pdf conditioned on mode $s \in \{1, 2\}$ at slot t is obtained by fusing the contributions from modes 1 and 2 at slot $t-1$ (denoted by the green lines in the figure). After receiving new observation $\mathbf{z}_t$, each mode updates the first two moments of the state pdf. Aided by $l_t^s = \mathcal{N}(\mathbf{z}_t; \hat{\mathbf{z}}_{t|t-1}^s, \mathbf{\Phi}_{t|t-1}^s)$, each mode obtains the posterior weight, based on which the fused state moments in the green box are acquired.

Thus, the posterior mean (27a) is the minimum mean-square error (MMSE) estimator of $\mathbf{x}_t$, whose uncertainty is characterized by the covariance matrix (27b). On the other hand, upon approximating the GM in (26) with a single Gaussian pdf having matched moments, (27a) can also be interpreted as the MAP estimator of $\mathbf{x}_t$.

The implementation steps of the IMGM algorithm for one recursion are summarized in Alg.1, and the flowchart of IMGM for $S = 2$ modes is presented in Fig. 1. Note that at initialization, the mode probabilities and mode-conditional state pdfs are set to be identical across modes; that is,

$$w_{0|0}^s = \frac{1}{S}, \quad \hat{\mathbf{x}}_{0|0}^s = \hat{\mathbf{x}}_0, \quad \mathbf{P}_{0|0}^s = \mathbf{P}_0, \quad s = 1, \ldots, S \tag{28}$$

where $\hat{\mathbf{x}}_0$ and $\mathbf{P}_0$ encode our prior information about the initial state distribution.

IMGM incurs low computational complexity of order $\mathcal{O}(STN^3)$ over T slots, which is clearly more affordable than the exponential complexity of the optimal solution of (9). To further maintain scalability for $N \gg$, the graph can be divided into N_g subgraphs, each with at most $\lceil N/N_g \rceil$ nodes. Upon leveraging distributed solvers along the lines of [28], the computational complexity per subgraph is $\mathcal{O}(S\lceil N/N_g \rceil^3)$, yielding an overall complexity of order $\mathcal{O}(SN_g\lceil N/N_g \rceil^3)$. Hence, scalability for large graphs can be effected by adjusting N_g. However, how to optimally choose N_g and divide the graph based on the topology, is an interesting future direction.

A few remarks are now in order.

Remark 1: IMGM is a memoryless online algorithm that requires no storage of past observations. All information about the past is summarized by the parameter set $\mathcal{P}_{t-1}$ that defines the GM pdf for the marginal state distribution.

Remark 2: Different from our IMGM, the classical IMM [6] first approximates a GM by a single Gaussian for each mode that corresponds to an updated mode-conditional state posterior, which is the input to one of S parallel mode-dependent Kalman filters with prediction and correction. Adhering to both Gaussian predicted (19) and posterior (22) mode-conditional state pdfs, IMGM predicts a GM per mode (17) that is then approximated by a single mode-conditional Gaussian (19), before running S parallel Kalman correction steps. The order of approximation and prediction makes no difference for linear state transition models.

IV. Multi-Kernel, Transition, and Noise Adaptivity

This section shows that IMGM can also be utilized to track dynamic processes and adapt the graph kernel(s), transition function, and noise variance per slot even for a fixed graph.

To start with, the linear state transition model in (6) may not be able to fully capture the dynamics of the graph processes,

Authorized licensed use limited to: University of Minnesota. Downloaded on July 09,2021 at 01:19:13 UTC from IEEE Xplore. Restrictions apply.

necessitating a general nonlinear transition model, which is given by a nonlinear function $f(\mathbf{A}^{\sigma_t}, \mathbf{x}_{t-1})$. Moreover, the state in several applications may adhere to a different dynamic model per slot. For example, the stocks in an economic network abide by different evolution patterns, e.g. in a period of economic recession. Hence, the transition function f (7) may jump among a candidate set of L transition functions $\{f^1, \ldots, f^L\}$ for a given topology at slot t.

Besides f, the noise model in (6) can be sensitive to the selection of the appropriate kernel (4). To deal with this, a dictionary of candidate 'basis kernels' (4) can be constructed with energy mappings in the set $\{r^1(\cdot) \ldots r^K(\cdot)\}$. Hence, the extended dynamical model of graph processes that further accounts for multiple kernels and switching nonlinear dynamic functions is given by

$$\mathbf{x}_t = f^{l_t}(\mathbf{A}_t^{\sigma_t}, \mathbf{x}_{t-1}) + \boldsymbol{\eta}_t^{k_t,\sigma_t} \tag{29}$$

where $l_t \in \mathcal{L} := \{1, \ldots, L\}$ is the active dynamic function index, and $\sigma_t \in \mathcal{S}$ denotes the active topology index; while the zero-mean Gaussian process noise $\boldsymbol{\eta}_t^{k_t,\sigma_t}$ has covariance matrix $\mathbf{K}_t^{k_t,\sigma_t}$ with active kernel function index $k_t \in \mathcal{K} := \{1, \ldots, K\}$.

On the other hand, the observation noise covariance in (1) is selected as $\mathbf{R}_t = \mu \mathbf{I}_M$. The scale μ is typically tuned via cross-validation offline among a candidate set of grid points $\{\mu^1, \ldots, \mu^R\}$. However, μ remains fixed for all t, and does not adapt to the data across slots t. To avoid cross-validation and effect a data-driven choice of μ, we can recast the observation model in (1) as

$$\mathbf{z}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{e}_t^{r_t} \tag{30}$$

where the covariance matrix of $\mathbf{e}_t^{r_t}$ is $\mathbf{R}_t^{r_t} = \mu^{r_t} \mathbf{I}_M$ with $r_t \in \mathcal{R} := \{1, \ldots, R\}$.

To incorporate switching topologies, dynamic functions, kernels, and observation noise covariances, we construct $\bar{S} = S \times L \times K \times R$ candidate dynamical models with the active model $\{f^{l_t}(\mathbf{A}^{\sigma_t}, \cdot), \mathbf{K}_t^{k_t,\sigma_t}, \mathbf{R}_t^{r_t}\}$ indicated by the extended network mode $\boldsymbol{\sigma}_t := (l_t, \sigma_t, k_t, r_t) \in \bar{\mathcal{S}}$, where the extended network mode set is $\bar{\mathcal{S}} := \mathcal{L} \times \mathcal{S} \times \mathcal{K} \times \mathcal{R}$. Before invoking the IMGM algorithm, one has to rescale the cost function in (9) by replacing $\mathbf{R}_t^{r_t}$ with $\mathbf{I}_M$, and subsequently absorbing $\mathbf{R}_t^{r_t}$ into the process noise covariance as $\tilde{\mathbf{K}}_t^{k_t,\sigma_t,r_t} = \mathbf{K}_t^{k_t,\sigma_t}/\mu^{r_t}$, such that the fitting error in (9) will have no scaling factor. The expanded candidate list of dynamical models at slot t is then constructed as $\{f^l(\mathbf{A}^\sigma, \cdot), \tilde{\mathbf{K}}_t^{k,\sigma,r}, \boldsymbol{\sigma} \in \bar{\mathcal{S}}\}$. Subsequently, by changing $\mathbf{R}_t$ to $\mathbf{I}_M$ in (23b), the IMGM algorithm is readily applied with only one revision in (15) for nonlinear dynamical models. As alluded to in the previous discussion, IMGM strives to maintain a Gaussian mode-conditional state pdf. Thus, to approximate the nonlinear transformation of a Gaussian state pdf by another Gaussian, we can leverage the unscented transformation as in unscented KF [34], or, just linearize the nonlinear transition functions, as in extended KF [6].

Three more remarks are in order.

Remark 3: For dynamical models with unknown parameters, candidate dynamical models can be constructed and IMGM can learn the model parameters that best fit the data on-the-fly, thus circumventing extra offline model training. Such a joint system identification and state estimation problem has been considered in the KF literature along three prevailing lines. The first leverages the expectation-maximization algorithm to iterate between state estimation and system identification, but the online characteristic is compromised; see e.g., [33]. The second approach chooses the model parameters from a known dictionary, and applies the classical IMM approach to select the appropriate parameters online [18]. Recently, for models with unknown process, and observation noise covariance matrices, a variational Bayesian approach is employed to obtain pdf estimates [13].

Remark 4: With only one graph and $\mathbf{F}_t^{l_t,\sigma_t} = \mathbf{0}$, IMGM offers an online probabilistic multi-kernel based alternative to reconstruct TI graph processes, which complements rather nicely the deterministic multi-kernel KRR framework in [26]. For a fixed set of nodes, the complexity of IMGM for time-invariant graphs is the same as that for the time-varying case.

V. Numerical Tests

In this section, we evaluate the performance of the proposed IMGM approach using synthetic and real data, and compare it with existing algorithms, including the kernel Kalman filter (KKF) [25]; the adaptive least mean-square (LMS) algorithm [10] with bandwidth $B_{\text{LMS}} \in \{2, 4, 6, \ldots, 20\}$ and step size $\mu_{\text{LMS}} \in \{0.5, 0.6, 0.7, \ldots, 2\}$; as well as the distributed least-squares reconstruction (DLSR) [35] with bandwidth $B_{\text{DLSR}} \in \{2, 4, 6, \ldots, 20\}$ and step sizes $\mu_{\text{DLSR}} \in \{0.2, 0.4, 0.6, \ldots, 2\}$ and $\beta_{\text{DLSR}} \in \{0.1, 0.2, \ldots, 0.9\}$. Both LMS and DLSR can track slowly time-varying B-bandlimited graph processes. Unless stated otherwise, the reported performances of LMS and DLSR are best-performing in terms of NMSE with hyperparameters selected from the candidate sets. Also, we consider the oracle of IMGM (abbreviated as "IMGM-O"), which relies on the dynamical model (6), but with known σ_t. To compare on equal footing with LMS and DLSR, which cannot deal with time-varying observation matrices, we set $\mathbf{H}_t = \mathbf{H}$ for all $t \in \{1, \ldots, T\}$. For experiments with switching graphs, the competing algorithms know the active graph topology per slot t, whereas our mode-agnostic IMGM estimates σ_t on-the-fly. The performance metric is the normalized mean-square error (NMSE) over unobserved nodes, which is given by

$$\text{NMSE}(t) := \|\mathbf{H}_t^c \left(\hat{\mathbf{x}}_{t|t} - \mathbf{x}_t\right)\|_2^2 / \|\mathbf{H}_t^c \mathbf{x}_t\|_2^2 \tag{31}$$

where $\mathbf{H}_t^c$ is the sampling matrix for the unobserved nodes. Due to the random sampling scheme, the performance is averaged over 100 random sampling realizations.

A. Synthetic Data

We consider a synthetic dynamic process over a network with $N = 60$ nodes, and $S = 2$ modes. The graph topologies associated with the two modes at slot 1 are generated by two symmetric Erdős-Rényi random graphs with edge existence probabilities 0.1 and 0.2, respectively. In the following slots, we randomly choose two pairs of nodes, and the edge between each pair is flipped relative to the previous slot per mode. The network

Authorized licensed use limited to: University of Minnesota. Downloaded on July 09,2021 at 01:19:13 UTC from IEEE Xplore. Restrictions apply.

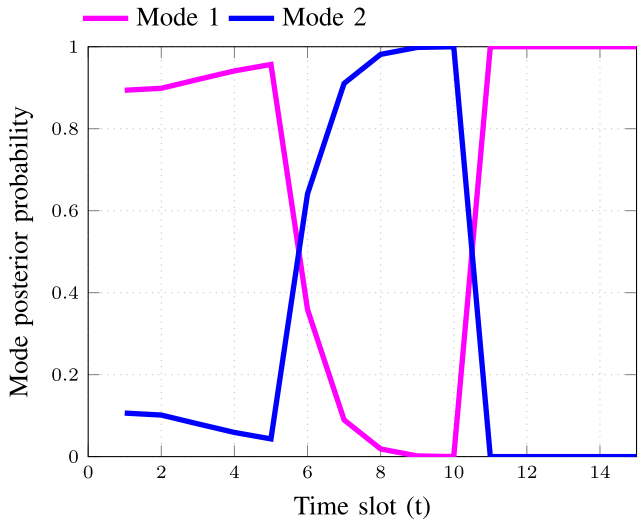


Fig. 2. Posterior mode pmfs of IMGM for synthetic data.

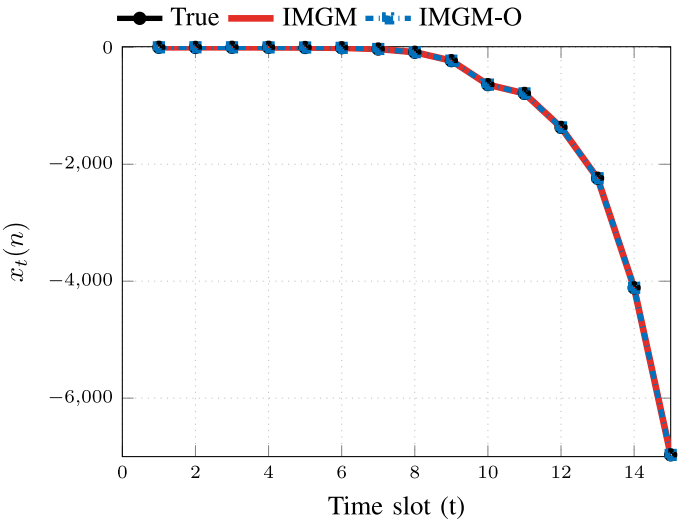


Fig. 4. True and estimated processes over an unobserved node for synthetic data.

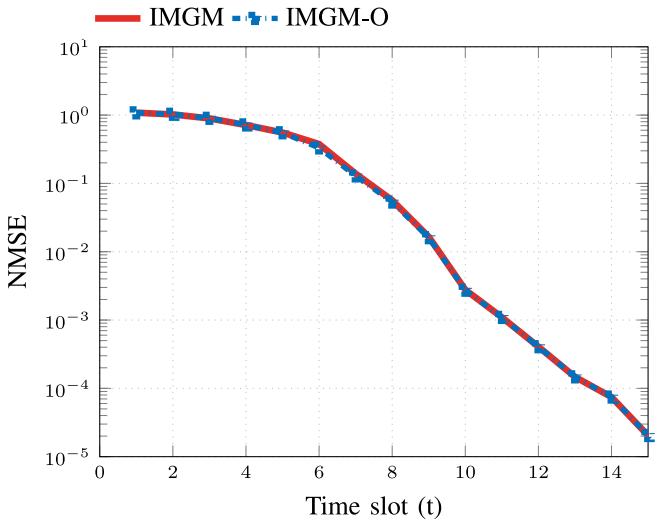


Fig. 3. NMSE for synthetic data.

switches from mode 1 to mode 2 at slot 6, and back to mode 1 at slot 11 over a total $T = 15$ slots. The dynamic graph process $\mathbf{x}_t$ is generated according to (6) with $\mathbf{F}_t^{\sigma_t} = 0.2(\mathbf{A}_t^{\sigma_t} + \mathbf{I}_N)$ and $\boldsymbol{\eta}_t^{\sigma_t} \sim \mathcal{N}(\boldsymbol{\eta}_t^{\sigma_t}; \mathbf{0}, \mathbf{K}_t^{\sigma_t})$, where $\mathbf{K}_t^{\sigma_t}$ is a diffusion kernel with $a = 0.1$ (see). The observations are generated based on (1) with $M = 30$ and $\mathbf{R} = 4\mathbf{I}_M$. Only IMGM-O was compared with IMGM because the rest of the approaches have no information about the generative model. The average mode posterior probabilities produced by IMGM over 100 Monte-Carlo runs are shown Fig. 2, which demonstrates that the IMGM is capable of keeping track of the active network mode in the presence of unknown switches. Further, Fig. 3 plots the NMSE over time, and illustrates that IMGM achieves the same NMSE as IMGM-O, which relies on extra information. Fig. 4 depicts the estimated processes along with the corresponding true values over an unobserved node. The perfect tracking of the true signal further validates IMGM's nearly optimal reconstruction performance.

B. Brain ECoG Dataset

Next, we experiment with the brain ECoG data obtained from an epilepsy study [17]. The ECoG time series were obtained from $N = 76$ electrodes implanted in a patient's brain before and after a seizure, where the onset of the seizure was identified by a neurophysiologist. Therefore, there are $S = 2$ modes, the pre-ictal and ictal mode that correspond to before and after the seizure. We extract 250 samples from the dataset for each of the two modes, which are preprocessed by subtracting the sample mean and normalizing by the sample standard deviation. The preprocessed samples are then concatenated so that $\sigma_t = 1$ for $t = 1, \ldots, 250$, and $\sigma_t = 2$ for $t = 251, \ldots, 500$. We construct a time-invariant symmetric correlation graph for each of the two modes, which is a special case of the problem statement at the end of Section II. The ECoG signals are modeled to evolve based on (6), where the state transition matrix $\mathbf{F}_t^{\sigma_t} = 0.15(\mathbf{A}^{\sigma_t} + \mathbf{I}_N)$, and process noise covariance $\mathbf{K}_t^{\sigma_t}$ is a diffusion kernel with parameter $a = 2$. Here, the value 0.15 and $a = 2$ are selected to yield the lowest NMSE from the sets $\{0.1, 0.11, 0.12, \ldots, 0.3\}$ and $\{1, 1.2, 1.4, \ldots, 3\}$. The observations are generated as in (1) with $M = 53$, and $\mathbf{R} = 10^{-2}\mathbf{I}_M$.

Fig. 5 shows the posterior mode probabilities $\{w_{t|t}^s\}_{s=1}^2$ produced by IMGM over 100 random sampling schemes. Here, IMGM plays the role of a "neurophysiologist" who detects the onset of an epileptic seizure. In addition, the NMSE of IMGM is comparable to that of the mode-clairvoyant IMGM-O, while markedly outperforming KKF, LMS and DLSR, as confirmed by Fig. 6. The NMSEs for all methods undergo a peak at the onset of the ictal, while for LMS and KKF the NMSEs are considerably larger during the ictal period. As in Fig. 7, the estimated brain signals from IMGM and IMGM-O over an unobserved node agree quite well with the corresponding true values, which is however not the case for the rest of the approaches. Further, Fig. 8 demonstrates that IMGM enjoys lower NMSE as the number M of sampled nodes grows.

Authorized licensed use limited to: University of Minnesota. Downloaded on July 09,2021 at 01:19:13 UTC from IEEE Xplore. Restrictions apply.

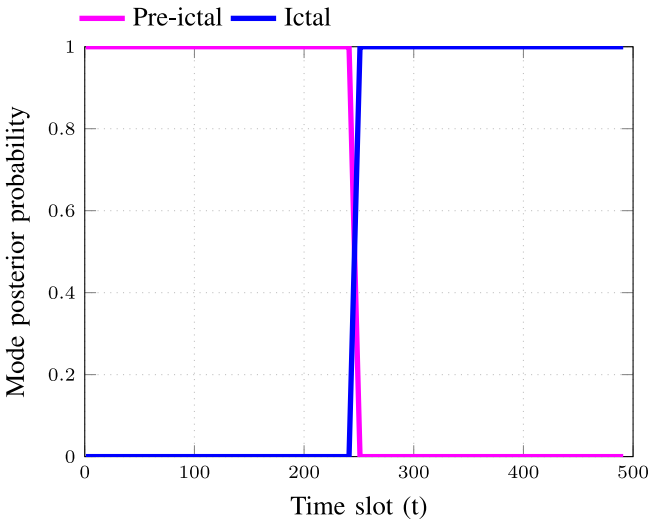


Fig. 5. Posterior mode pmfs of IMGM for ECoG data.

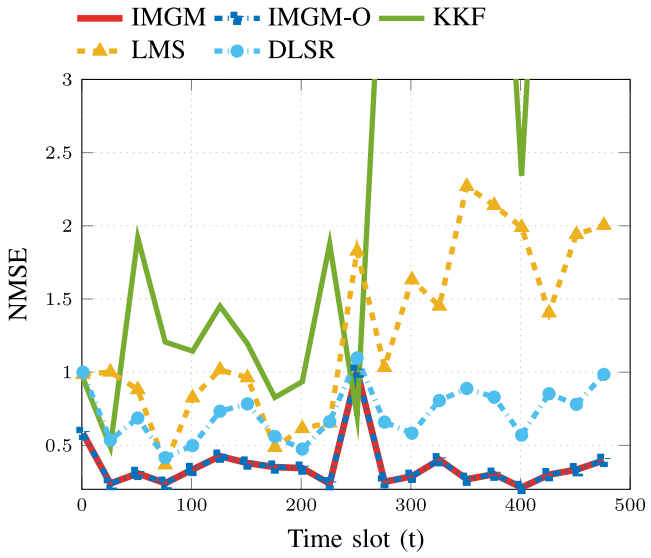


Fig. 6. NMSE for ECoG data ($\mu_{\text{LMS}} = 0.6$, $B_{\text{LMS}} = 2$, $\mu_{\text{DLSR}} = 1.2$, $B_{\text{DLSR}} = 6$, $\beta_{\text{DLSR}} = 0.5$).

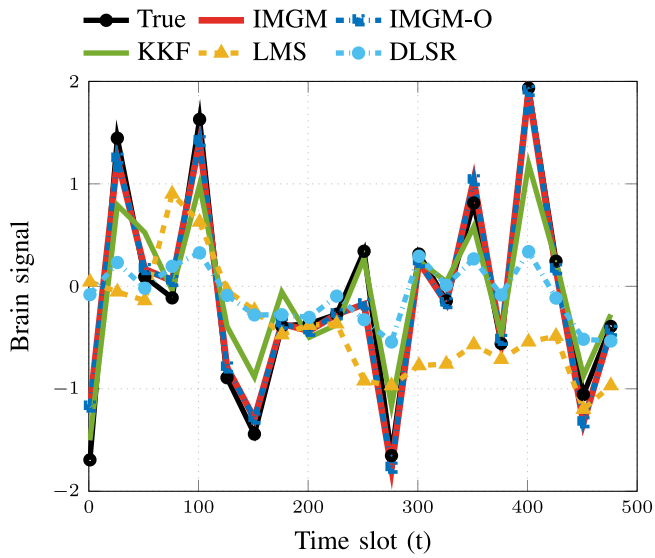


Fig. 7. True and estimated brain signals over an unobserved node for ECoG data ($\mu_{\text{LMS}} = 0.6$, $B_{\text{LMS}} = 2$, $\mu_{\text{DLSR}} = 1.2$, $B_{\text{DLSR}} = 6$, $\beta_{\text{DLSR}} = 0.5$).

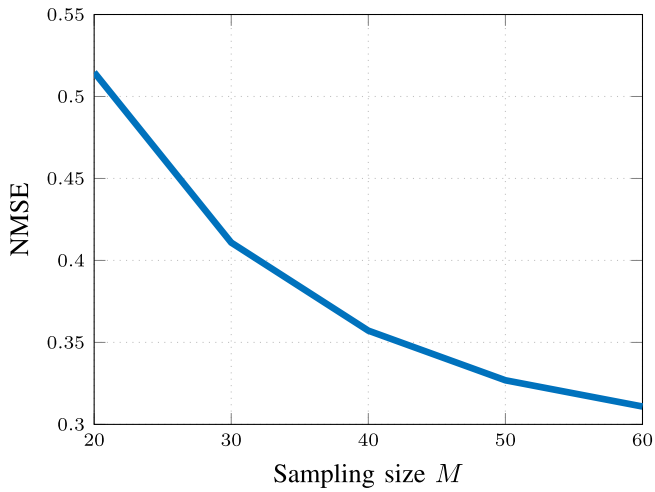


Fig. 8. Overall NMSE of IMGM versus M for ECoG data.

C. Temperature Prediction

The next dataset comprises hourly temperature measurements at $N = 109$ measuring stations across the continental United States in 2010 by the National Climatic Data Center [1]. A time-invariant graph was constructed based on geographical distances as in [25]. Even though only one graph is available, IMGM can still be applied to track the dynamic processes and simultaneously learn the model parameters that best fit the data as in Section IV, which would otherwise need an offline training process. The value $x_t(n)$ represents the tth temperature sample recorded at the nth station. The sampling interval in our experiment is chosen to be one day. The number of observed nodes is $M = 44$, and observation noise covariance is selected from the candidate set as $\mathbf{R}^{r_t} = \mu^{r_t}\mathbf{I}_M$, where $\mu^{r_t} \in 10^{-4} \times \{1, 2, \ldots, 5\}$. The transition matrix is taken as $\mathbf{F}_t^{l_t} = c^{l_t}(\mathbf{A} + \mathbf{I}_N)$, where c^{l_t} takes value from 0.05 to 0.15 with uniform grid 0.02. The process noise is given by a diffusion kernel with $a = 2$. Thus, IMGM is equipped with 30 candidate dynamical models, among which the best performing one is assigned to IMGM-O with $\mathbf{F}_t = 0.05(\mathbf{A} + \mathbf{I}_N)$ and $\mathbf{R} = 10^{-4}\mathbf{I}_M$.

As shown in Fig. 9, the mode-agnostic IMGM demonstrates superior tracking performance compared to the KKF, LMS and DLSR, while it also showcases performance comparable to IMGM-O. Hence, IMGM is capable of selecting the dynamical model that best fits the data on-the-fly. Fig. 10 further corroborates this assertion by displaying the true and estimated network delays from the candidate approaches over an unobserved node. The probability of existence of each model is reported by the posterior mode pmf $w_{t|t}^s$ as $w_{t|t}^1 \approx 1$, and $w_{t|t}^s \approx 0$ for s otherwise.

Authorized licensed use limited to: University of Minnesota. Downloaded on July 09,2021 at 01:19:13 UTC from IEEE Xplore. Restrictions apply.

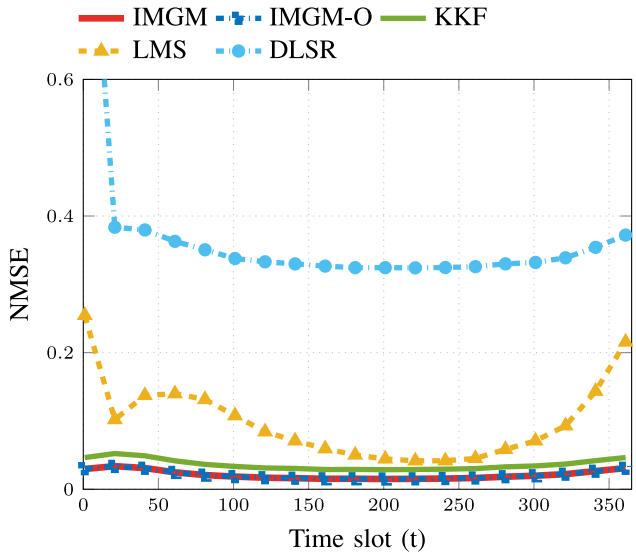


Fig. 9. NMSE for temperature data ($\mu_{\text{LMS}} = 1.5$, $B_{\text{LMS}} = 10$, $\mu_{\text{DLSR}} = 1.2$, $B_{\text{DLSR}} = 4$, $\beta_{\text{DLSR}} = 0.5$).

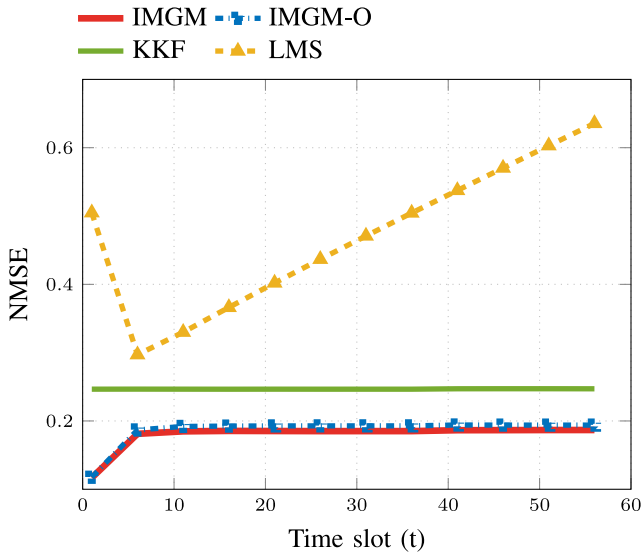


Fig. 11. NMSE for network delay data ($\mu_{\text{LMS}} = 1.5$, $B_{\text{LMS}} = 12$).

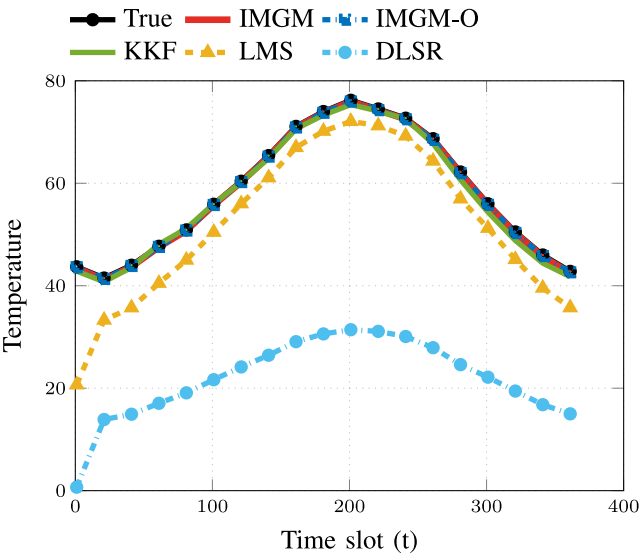


Fig. 10. True and estimated temperature values over an unobserved location ($\mu_{\text{LMS}} = 1.5$, $B_{\text{LMS}} = 10$, $\mu_{\text{DLSR}} = 1.2$, $B_{\text{DLSR}} = 4$, $\beta_{\text{DLSR}} = 0.5$).

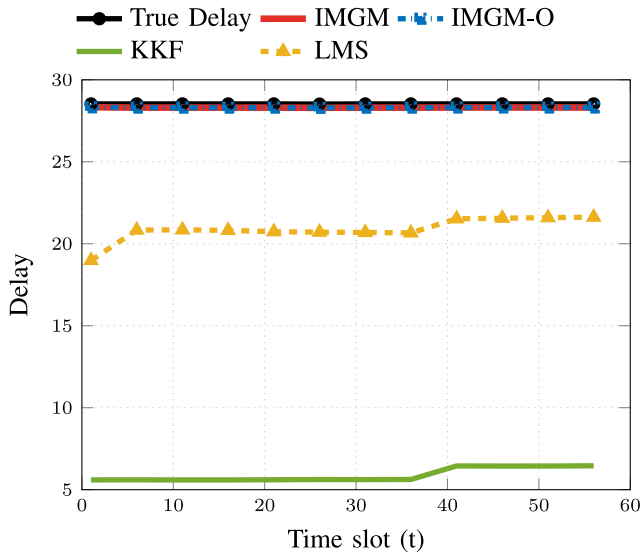


Fig. 12. True and estimated network delays over an unobserved path ($\mu_{\text{LMS}} = 1.5$, $B_{\text{LMS}} = 12$).

D. Network Delay Prediction

The last dataset records measurements of path delays on the Internet2backbone [2]. The network comprises 9 end-nodes and 26 directed links. There are $N = 70$ paths, each connecting two origin-destination nodes by a subset of the 26 links. The active links for each path are described by the path-link routing matrix $\mathbf{B} \in \{0, 1\}^{70\times 26}$, whose (n, l)th entry $B_{n,l}$ is 1, if path n traverses link l, and 0 otherwise. With each vertex representing one of these paths, an undirected graph is constructed with the (n, n')th entry $(n \neq n')$ of the adjacency matrix as

$$A(n, n') = \frac{\sum_{l=1}^{26} B_{n,l} B_{n',l}}{\sum_{l=1}^{26} B_{n,l} + \sum_{l=1}^{26} B_{n',l} - \sum_{l=1}^{26} B_{n,l} B_{n',l}}$$

which places large weights for vertices (paths) with a large number of common links. The graph process $x_t(n)$ represents the delay of path n in minutes.

The number of observed nodes is selected to be $M = 20$. The candidate dynamical models for IMGM are configured as follows. The state transition matrix is selected to be $\mathbf{F}_t = 0.17(\mathbf{A} + \mathbf{I}_N)$. Process noise covariance $\mathbf{K}^{k_t}$ is chosen from a set of $K = 8$ diffusion kernels (cf.) with parameter a^{k_t} taking values from 0.6 to 2 with uniform space 0.2. Observation noise covariance for the observation model (30) is $\mathbf{R}^{r_t} = \mu^{r_t}\mathbf{I}_M$, with $\mu^{r_t} \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$. The number of candidate models for IMGM is then $\bar{S} = 8 \times 4 = 32$, among which IMGM-O is equipped with the best performing one: $a = 0.6$ and

Authorized licensed use limited to: University of Minnesota. Downloaded on July 09,2021 at 01:19:13 UTC from IEEE Xplore. Restrictions apply.

$\mathbf{R} = 10^{-2}\mathbf{I}_M$. For this experiment, we did not employ DLSR because it did not yield comparable performance to the rest of the alternatives.

Adaptively choosing a model with kernel parameter a^{k_t} and noise parameter μ^{r_t} from the candidate set, IMGM exhibits superior tracking performance compared to the single-model alternatives, as confirmed by Fig. 11. This also corroborates that IMGM provides a probabilistic multi-kernel learning approach. The estimated and true network delays from an unobserved node over the entire observation interval are plotted in Fig. 12.

VI. Conclusion

This paper dealt with tracking dynamic graph processes that evolve over a candidate set of graph topologies with unknown switches. To this end, a dynamical model was introduced to capture both spatial and temporal variations of the graph processes through the notion of active mode-conditioned topology. Subsequently, given observations over a subset of nodes, a scalable Bayesian tracker, termed IMGM, was developed to carry out semi-supervised tracking of the dynamic graph processes jointly with the active network mode. The novel IMGM solver lends itself to several important generalizations, including dynamic function switches, multiple kernels, and adaptive observation noise covariances. Accounting for all these models, IMGM offers an online approach to select the one that best fits the data adaptively, while at the same time tracks the graph processes. Numerical tests on synthetic and real data corroborated the performance gain of the novel IMGM approach.

References

[1] "1981-2010 U.S. climate normals," [Online]. Available: https://www. ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/climat e-normals/1981-2010-normals-data. Accessed: Apr. 29, 2019.

[2] "One-way ping internet2," [Online]. Available: http://software.internet2.edu/owamp/, Accessed: Apr. 29, 2019.

[3] R. Agaev and P. Chebotarev, "The matrix of maximum out forests of a digraph and its applications," *Autom. Remote Control*, vol. 61, no. 9, pp. 1424–1450, 2000.

[4] R. P. Agaev and P. Chebotarev, "Spanning forests of a digraph and their applications," *Autom. Remote Control*, vol. 62, no. 3, pp. 443–466, Mar. 2001.

[5] B. Baingana and G. B. Giannakis, "Tracking switched dynamic network topologies from information cascades," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 985–997, Feb. 2017.

[6] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation With Applications to Tracking and Navigation: Theory Algorithms and Software*. Hoboken, NJ, USA: Wiley, 2004.

[7] H. A. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Trans. Autom. Control*, vol. 33, no. 8, pp. 780–783, Aug. 1988.

[8] D. Boley, G. Ranjan, and Z.-L. Zhang, "Commute times for a directed graph using an asymmetric Laplacian," *Linear Algebra its Appl.*, vol. 435, no. 2, pp. 224–242, Jul. 2011.

[9] F. Chung, "Laplacians and the Cheeger inequality for directed graphs," *Ann. Combinatorics*, vol. 9, no. 1, pp. 1–19, Apr. 2005.

[10] P. Di Lorenzo, S. Barbarossa, P. Banelli, and S. Sardellitti, "Adaptive least mean-squares estimation of graph signals," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 2, no. 4, pp. 555–568, Dec. 2016.

[11] P. A. Forero, K. Rajawat, and G. B. Giannakis, "Prediction of partially observed dynamical processes over networks via dictionary learning," *IEEE Trans. Signal Process.*, vol. 62, no. 13, pp. 3305–3320, Jul. 2014.

[12] G. B. Giannakis, Y. Shen, and G. V. Karanikolas, "Topology identification and learning over graphs: Accounting for nonlinearities and dynamics," *Proc. IEEE*, vol. 106, no. 5, pp. 787–807, May 2018.

[13] Y. Huang, Y. Zhang, Z. Wu, N. Li, and J. Chambers, "A novel adaptive Kalman filter with inaccurate process and measurement noise covariance matrices," *IEEE Trans. Autom. Control*, vol. 63, no. 2, pp. 594–601, Feb. 2018.

[14] V. N. Ioannidis, D. Romero, and G. B. Giannakis, "Inference of spatio-temporal functions over graphs via multikernel kriged Kalman filtering," *IEEE Trans. Signal Process.*, vol. 66, no. 12, pp. 3228–3239, Jun. 2018.

[15] E. D. Kolaczyk, *Statistical Analysis of Network Data: Methods and Models*. Berlin, Germnay: Springer, 2009.

[16] R. I. Kondor and J. Lafferty, "Diffusion kernels on graphs and other discrete structures," in *Proc. Int. Conf. Machi. Learn.*, Jul. 2002, pp. 315–322.

[17] M. A. Kramer, E. D. Kolaczyk, and H. E. Kirsch, "Emergent network topology at seizure onset in humans," *Epilepsy Res.*, vol. 79, no. 2/3, pp. 173–186, May 2008.

[18] X. R. Li and Y. Bar-Shalom, "A recursive multiple model approach to noise identification," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 3, pp. 671–684, Jul. 1994.

[19] X.-R. Li and Y. Bar-Shalom, "Design of an interacting multiple model algorithm for air traffic control tracking," *IEEE Trans. Control Syst. Technol.*, vol. 1, no. 3, pp. 186–194, Sep. 1993.

[20] Q. Lu, V. Ioannidis, and G. B. Giannakis, "Semi-supervised tracking of dynamic processes over switching graphs," in *Proc. IEEE Data Sci. Workshop*, Minneapolis, MN, Jun. 2019, pp. 64–68.

[21] Q. Lu, V. Ioannidis, G. B. Giannakis, and M. Coutino, "Learning graph processes with multiple dynamical models," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2019, pp. 1783–1787.

[22] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, "Interacting multiple model methods in target tracking: A survey," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 1, pp. 103–123, Jan. 1998.

[23] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.

[24] N. Perraudin and P. Vandergheynst, "Stationary signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3462–3477, Jul. 2017.

[25] D. Romero, V. N. Ioannidis, and G. B. Giannakis, "Kernel-based reconstruction of space-time functions on dynamic graphs," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 6, pp. 856–869, Sep. 2017.

[26] D. Romero, M. Ma, and G. B. Giannakis, "Kernel-based reconstruction of graph signals," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 764–778, Feb. 2017.

[27] A. V. Savkin and R. J. Evans, *Hybrid Dynamical Systems: Controller and Sensor Switching Problems*. Berlin, Germany: Springer, 2002.

[28] I. D. Schizas, G. B. Giannakis, S. I. Roumeliotis, and A. Ribeiro, "Consensus in Ad Hoc WSNs with noisy links—part II: Distributed estimation and smoothing of random signals," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1650–1666, Apr. 2008.

[29] S. Segarra, A. G. Marques, G. Leus, and A. Ribeiro, "Reconstruction of graph signals through percolation from seeding nodes," *IEEE Trans. Signal Process.*, vol. 64, no. 16, pp. 4363–4378, Aug. 2016.

[30] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.

[31] A. J. Smola and R. Kondor, "Kernels and regularization on graphs," in *Learning Theory and Kernel Machines*. Berlin, Germany: Springer, 2003, pp. 144–158.

[32] P. A. Traganitis and G. B. Giannakis, "Sketched subspace clustering," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1663–1675, Apr. 2018.

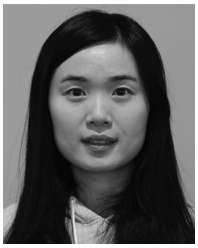
[33] E. A. Wan and A. T. Nelson, "Dual Kalman filtering methods for nonlinear prediction, smoothing and estimation," in *Proc. Advances Neural Inf. Process. Syst.*, 1997, pp. 793–799.

[34] E. A. Wan and R. Van Der Merwe, "The unscented Kalman filter for nonlinear estimation," in *Proc. IEEE Adaptive Syst. for Signal Process., Commun., Control Symp.*, 2000, pp. 153–158.

[35] X. Wang, M. Wang, and Y. Gu, "A distributed tracking algorithm for reconstruction of graph signals," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 4, pp. 728–740, Jun. 2015.

[36] Q. Yu *et al.*, "Application of graph theory to assess static and dynamic brain connectivity: Approaches for building brain graphs," *Proc. IEEE*, vol. 106, no. 5, pp. 886–906, May 2018.

Authorized licensed use limited to: University of Minnesota. Downloaded on July 09,2021 at 01:19:13 UTC from IEEE Xplore. Restrictions apply.



Qin Lu (Member, IEEE) received the Ph.D. degree in electrical engineering from the University of Connecticut (UConn), Mansfield, CT, USA, in 2018. She is currently a Postdoctoral Research Associate with the University of Minnesota, Twin Cities, Minneapolis, MN, USA. Her current research interests include machine learning, data science, and network science. In the past, she has worked on statistical signal processing, multiple target tracking, and underwater acoustic communications. She was awarded Summer Fellowship and Doctoral Dissertation Fellowship at UConn. She was also the recipient of the Women of Innovation Award in Collegian Innovation and Leadership by Connecticut Technology Council in March, 2018.

Vassilis N. Ioannidis (Student Member, IEEE) received the diploma in electrical and computer engineering from the National Technical University of Athens, Athens, Greece, in 2015, and the M.Sc. degree in electrical engineering in 2017 from the University of Minnesota, Twin Cities, Minneapolis, MN, USA. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering. He received the Doctoral Dissertation Fellowship in 2019 from the University of Minnesota. He also received student travel awards from the IEEE Signal Processing Society in 2017–2018 and from the IEEE (2018). From 2014 to 2015, he was a middleware consultant for Oracle in Athens, Greece, and received a Performance Excellence Award. His research interests include deep graph learning, machine learning, big data analytics, and network science.

Georgios B. Giannakis (Fellow, IEEE) received the diploma in electrical engineering from the National Technical University of Athens, Athens, Greece, 1981, the M.Sc. degree in electrical engineering, the M.Sc. degree in mathematics, and the Ph.D. degree in electrical engineering from the University of Southern California (USC), Los Angeles, CA, USA, in 1983, 1986, and 1986, respectively. From 1982 to 1986, he was with the USC. He was a Faculty Member with the University of Virginia from 1987 to 1998, and since 1999, he has been a Professor with the University of Minnesota, Twin Cities, Minneapolis, MN, USA, where he holds an ADC Endowed Chair, a University of Minnesota McKnight Presidential Chair in ECE, and serves as the Director of the Digital Technology Center. His current research interests focus on data science, and network science with applications to the Internet of Things, and power networks with renewables. His general interests span the areas of statistical learning, communications, and networking—subjects on which he has authored or coauthored more than 460 journal papers, 760 conference papers, 25 book chapters, two edited books and two research monographs. He is the Coinventor of 33 issued patents, and the corecipient of 9 best journal paper awards from the IEEE Signal Processing (SP) and Communications Societies, including the G. Marconi Prize Paper Award in Wireless Communications. He also received the IEEE-SPS Nobert Wiener Society Award (2019); EURASIP's A. Papoulis Society Award (2020); Technical Achievement Awards from the IEEE-SPS (2000) and from EURASIP (2005); the IEEE ComSoc Education Award (2019); the G. W. Taylor Award for Distinguished Research from the University of Minnesota, and the IEEE Fourier Technical Field Award (2015). He is a fellow of the National Academy of Inventors, the European Academy of Sciences, and EURASIP. He has served the IEEE in a number of posts, including that of a Distinguished Lecturer for the IEEE-SPS.

Authorized licensed use limited to: University of Minnesota. Downloaded on July 09,2021 at 01:19:13 UTC from IEEE Xplore. Restrictions apply.